

The virion glycoproteins of Ebola viruses are encoded in two reading frames and are expressed through transcriptional editing

(filovirus/glycoprotein gene/phylogenetic analysis)

ANTHONY SANCHEZ*, SAM G. TRAPPIER, BRIAN W. J. MAHY, CLARENCE J. PETERS, AND STUART T. NICHOL

Special Pathogens Branch, Division of Viral and Rickettsial Diseases, National Center for Infectious Diseases, Centers for Disease Control and Prevention, Atlanta, GA 30333

Communicated by Kenneth I. Berns, Cornell University Medical College, New York, NY, December 27, 1995 (received for review October 12, 1995)

ABSTRACT In late 1994 and early 1995, Ebola (EBO) virus dramatically reemerged in Africa, causing human disease in the Ivory Coast and Zaire. Analysis of the entire glycoprotein genes of these viruses and those of other EBO virus subtypes has shown that the virion glycoprotein (130 kDa) is encoded in two reading frames, which are linked by transcriptional editing. This editing results in the addition of an extra nontemplated adenosine within a run of seven adenosines near the middle of the coding region. The primary gene product is a smaller (50–70 kDa), nonstructural, secreted glycoprotein, which is produced in large amounts and has an unknown function. Phylogenetic analysis indicates that EBO virus subtypes are genetically diverse and that the recent Ivory Coast isolate represents a new (fourth) subtype of EBO virus. In contrast, the EBO virus isolate from the 1995 outbreak in Kikwit, Zaire, is virtually identical to the virus that caused a similar epidemic in Yambuku, Zaire, almost 20 years earlier. This genetic stability may indicate that EBO viruses have coevolved with their natural reservoirs and do not change appreciably in the wild.

The 1995 epidemic of Ebola (EBO) virus disease in Zaire (1), coupled with the discovery of an EBO virus in the Ivory Coast in late 1994 (2), has sparked new scientific and public interest in these mysterious and highly pathogenic viruses. EBO viruses are nonsegmented negative-strand RNA viruses that are genetically related to, but distinct from, the Marburg (MBG) virus (3). These viruses, classified in the genus *Filovirus* in the family *Filoviridae*, cause a severe hemorrhagic disease in human and nonhuman primates (4, 5). The EBO group of filoviruses comprises three discrete subtypes: Zaire, Sudan, and Reston (EBO-Z, EBO-S, and EBO-R, respectively). In late November 1994, EBO virus was discovered in the Tai Forest of the Ivory Coast (1). This virus (EBO-IC) caused a single nonfatal human infection and was believed to have caused increased mortality in a troop of wild chimpanzees. Preliminary serological analysis of the EBO-IC isolate suggests that it may represent a fourth subtype (ref. 2; Centers for Disease Control and Prevention, unpublished data). The much publicized outbreak of human disease in the city of Kikwit, Zaire, in late Spring, 1995, involved over 300 cases with a 77% mortality rate (1, 6). This epidemic was caused by a strain of EBO-Z, the most pathogenic subtype of EBO virus, which had not been reported for 18 years.

The virion surface glycoproteins (GPs) of filoviruses are multimers of a single structural GP; they are important in the binding of virions to cell receptors and virus entry into the cell cytoplasm (3). This protein is expressed from the GP gene, which in the genomes of filoviruses is positioned fourth (from the 3' end) of seven linearly arranged genes (3, 7). Filovirus GPs are highly glycosylated, containing both N-linked and

O-linked carbohydrates (3, 8–10) that contribute from one-third to one-half of their relatively large molecular weight ($M_r \approx 130$ –170 kDa).

To better define the relationship of EBO viruses to one another, we have investigated the structure and expression of their GP genes.[†] Here we describe an unusual organization of the GP genes of all EBO virus subtypes, present a phylogenetic profile for the family *Filoviridae*, and discuss the role of GP gene products in the pathogenesis of EBO virus disease.

MATERIALS AND METHODS

Viruses. The initial EBO-R isolate (R/USA/Reston/1989/119810 strain) was obtained from P. B. Jahrling (U.S. Army Medical Research Institute for Infectious Diseases, Fort Detrick, MD) and was isolated and passaged once in MA-104 cells, plaque purified three times, and passaged once in Vero E6 cells. Two 1992 EBO-R viruses [R/Philippines/Manila/1992/920084, isolated at Centers for Disease Control and Prevention (Atlanta), and R/Italy/Siena/1992/12552, obtained from D. Brown (Public Health Laboratory Service, Virus Reference Laboratory, London)] were passaged 2 to 3 times in Vero E6 cells and once in MA-104 cells. A 1976 EBO-Z virus (Z/Zaire/Yambuku/1976/057935 also referred to as the Mayinga isolate) was passaged once in Vero cells and once in Vero E6 cells. The passage histories of other EBO-Z viruses (Z/Zaire/Yambuku/1976/057878 and Z/Zaire/Tandala/1977/088296, also referred to as the Eckron and Bonduni isolates, respectively) and EBO-S viruses (S/Sudan/Maridi/1976/VCP2D11 and S/Sudan/Nzara/1979/015176, also known as the Boneface and Maleo isolates, respectively) are described elsewhere (11). The EBO-IC (IC/Ivory Coast/Tai Forest/1994) was obtained from B. Le Guenno (Institute Pasteur, Paris), and was passaged 2 to 5 times in Vero E6 cells. The EBO-Z virus from the 1995 epidemic in Kikwit (Z/Zaire/Kikwit/1995/9510621) was isolated at the Centers for Disease Control and Prevention in Vero E6 cells inoculated with a blood specimen from an acutely infected patient (who later died). However, all sequence data related to this strain were derived from RNA extracted from the same human blood specimen.

Viral RNA Purification, Amplification, and Cloning. Preparation of genomic RNA (vRNA) from purified EBO virions and the extraction of total infected cell RNA were performed as described (12). Syntheses of cDNA from the vRNA of EBO-R (Reston/1989) and EBO-S (Nzara/1979) were performed (13), followed by blunt-end ligation into the *Sma* I site

Abbreviations: EBO, Ebola; MBG, Marburg; TCF, tissue culture fluid; GP, glycoprotein; SGP, small/secreted glycoprotein precursor; RT-PCR, reverse transcription-PCR; PMSF, phenylmethylsulfonyl fluoride; ORF, open reading frame; RIP, radioimmunoprecipitation.

*To whom reprint requests should be addressed.

[†]The sequences reported in this paper have been deposited in the GenBank data base (accession nos. U23069, U23152, U23187, U23416, U23417, U28006, U28077, and U28134).

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

of pUC18 or pSP73/pSP72 vectors (Promega) and cloned as previously described (14). Sequencing of cloned plasmid preparations was performed using the dideoxy chain termination method (15). A manual radiolabeling method employing cloned bacteriophage T7 DNA polymerase (Sequenase) and an automated nonisotopic method (dye-terminator cycle sequencing; Perkin-Elmer) was used. Direct (automated) sequencing of DNA amplified from viral sequences by reverse transcription-PCR (RT-PCR) was used in sequencing the EBO-S strains, EBO-IC, the Kikwit/1995 strain of EBO-Z, and the 1992 EBO-R isolates. RT-PCR amplification of purified vRNA sequences was performed using GenAmp RNA PCR kits (Perkin-Elmer) or reactions were assembled using reagents and buffers obtained from commercial sources (Promega and Boehringer Mannheim).

Quantitation of Transcriptional Editing. Editing of GP gene transcripts was quantitated for EBO-R (Reston/1989) and EBO-Z (Yambuku/1976/057935) by RT-PCR amplification of GP sequences from crude infected cell RNA (mRNA) preparations. First-strand cDNA synthesis was performed by heating approximately 5 μ g of total infected cell RNA and 300 ng of the primer 5'-CCGGTACC(T)₃₅ at 65°C for 1 min and placing the mixture at 42°C; the remaining components were then added (25- μ l reaction volume). Primer extension with reverse transcriptase proceeded for 40 min, followed by adding 100 μ l of sterile water and boiling for 3 min. Approximately 5–10 μ l of this diluted reaction solution served as template for the amplification of viral sequences. For PCR amplification of the EBO-R GP gene editing region, the primers 5'-CCTTAGCAACAGTACAGGGAGAT and 5'-GCCGTCTCTTGCGG-GTCCTGG were used. For the EBO-Z GP gene, the primers 5'-CTGGATCCAACAACACAATGGGCGTTACAGGAAT and 5'-AATGGTACCGACTCGTGGAGATTGTGG were used to amplify the editing region. The EBO-R PCR product was cleaved with the restriction enzyme *Bam*HI and ligated into the *Bam*HI site of pUC18. The EBO-Z PCR product was cleaved with *Kpn*I followed by *Bam*HI, directionally ligated into pUC18, and isolated clones sequenced.

To determine if virion (genomic) RNA might contain eight or more uridines at the editing site, vRNA was extracted from a human blood specimen (from which the Kikwit/1995 virus was isolated) and used as template in a RT-PCR to amplify a DNA fragment containing the editing site. The primer 5'-GGACCCGTCTAGTGGCTACTATTC (plus-sense) was used to prime first-strand cDNA synthesis from vRNA template molecules, as above, then the reaction was boiled 3 min before its use in PCR amplification. The above primer was then used with the negative-sense primer 5'-CCAATGCATGAT-GCGACACTGCAGCTTCCC (nonviral sequences are underlined) to amplify a 482-bp fragment containing the editing region. The RT-PCR product was isolated by agarose gel electrophoresis and extracted from the gel using a commercial kit [QIAEX II, Qiagen Inc. (Chatsworth, CA)]. Isolated DNA was directly ligated into the plasmid pCRII using a TA prokaryotic cloning kit (Invitrogen). Cloned plasmid DNA was isolated and individual clones were sequenced.

In Vitro Expression of EBO-R GP Gene Products. *In vitro* expression of EBO-R (Reston/1989) GP gene sequences was performed by RT-PCR amplification of the GP open reading frame (ORF) from total infected-cell RNA using the primers 5'-ATACCCGGGGCCCCAAATTACCTATACAACA and 5'-TTTTCTAGAATATTAATCATTATTAAGAGA. Amplified products were digested with *Xma*I and *Xba*I and directionally ligated into pSP73, cloned, and full-length inserts containing seven or eight adenosines in the editing site were isolated (pSP73-RESGP-7A and pSP73-RESGP-8A, respectively). Plasmids were linearized with *Hind*III transcribed with SP6 RNA polymerase, DNase-treated, and run-off transcripts were *in vitro* translated using a rabbit reticulocyte lysate system (Promega).

Radioimmunoprecipitation (RIP) of EBO Virus GPs. Monolayers of MA-104 cells in 24-well panels were infected with EBO viruses (≥ 10 plaque-forming units per cell) or mock infected. Cells were incubated for 3 to 4 days, washed with Dulbecco's minimal essential medium deficient in methionine and/or cysteine containing 2% dialyzed fetal bovine serum and antibiotics, and 300 μ l of the same medium plus 20 μ Ci/ml [³⁵S]methionine/cysteine. Cells were incubated for 4 hr, then tissue culture fluid (TCF) was removed and the cells were washed as above. The cell monolayer and 900 μ l of harvested TCF were treated with 100 μ l 1% SDS for 1 min, followed by the addition of 900 μ l 1.1 \times TNE buffer (1 \times TNE = 10 mM Tris-HCl/150 mM NaCl/3 mM EDTA, pH 7.6) containing 1.1% Nonidet P-40, 0.55% sodium deoxycholate, and 1.1 mM phenylmethylsulfonyl fluoride (PMSF). Triton X-100 treatment was performed by the addition of 1 \times TNE containing 1% Triton X-100 and 1 mM PMSF to cells; TCF was made to 1% with Triton X-100 and to 1 mM PMSF. RIP assays were performed on 500 μ l of solutions as described (14). Antibodies used in RIP assays were: (i) an anti-EBO-R GP/small/secreted glycoprotein precursor (SGP) monoclonal antibody (high-titered mouse ascitic fluid; cross-reactive with GP and SGP) prepared essentially as described (16), and (ii) a high-titered polyclonal mouse anti-EBO-Z GP produced against a GP expressed by a recombinant baculovirus (10). Endoglycosidase digestion of RIP products was performed as described elsewhere (17).

Computer-Aided Sequence Analyses. Computer analyses of nucleic acid and predicted amino acid sequences for the GP genes of EBO subtypes were performed using the Genetics Computer Group (Madison, WI) Sequence Analysis Software Package (Version 7.3-AXP) run on a Digital AXP Alpha Workstation. The PILEUP program was used to align nucleotide sequences with a GapWeight setting of 5.0 and a GapLength-Weight setting of 0.5. Phylogenetic analysis of nucleotide alignments was performed using the PAUP software Version 3.1.1 (developed by D. L. Swofford) run on a Power Macintosh model 8100/110.

RESULTS

The sequences of the entire GP genes of EBO-Z, EBO-S, EBO-R, and EBO-IC were determined from cloned cDNA and RT-PCR amplified DNA products (see Fig. 1 legend for GenBank accession numbers). The EBO-IC strain was found to differ substantially from the other subtypes of EBO virus ($\leq 60\%$ identity) and represents a fourth subtype. The organization and specific features of these genes are shown in Fig. 1. The most unusual finding was that the GPs of all EBO subtypes are encoded in two frames and that in each case an SGP was predicted as the primary gene product.

The GP genes of all EBO viruses begin and end with conserved transcription start and stop (polyadenylation) sites (14). The only difference in GP gene organization seen in the four subtypes is found in EBO-R. For the other subtypes, the transcription stop site overlaps the transcription start sites of the downstream (VP30) gene and is centered on the common pentanucleotide sequence 3'-UAAUU found in all transcriptional signals (3'-CUACUUCUAAUUCUUUUU). In contrast, the GP gene stop site of EBO-R is separated from the downstream gene by a short intergenic region of three nucleotides that is followed by a conserved start site (3'-UAAU-UCUUUUU-GAA-UACUGCUUCUAAUU).

The SGP and GP share the same N-terminal ≈ 300 residues, but have unique C-terminal sequences (Fig. 2). Unlike that of the GP, the C terminus of the SGP lacks a transmembrane anchor sequence, which results in its secretion from infected cells. The unique C-terminal portion of the SGP is hydrophilic and rich in charged residues (mostly positive). The entire SGP sequence for EBO subtypes is relatively conserved, but the central area of the GP contains a large variable hydrophilic

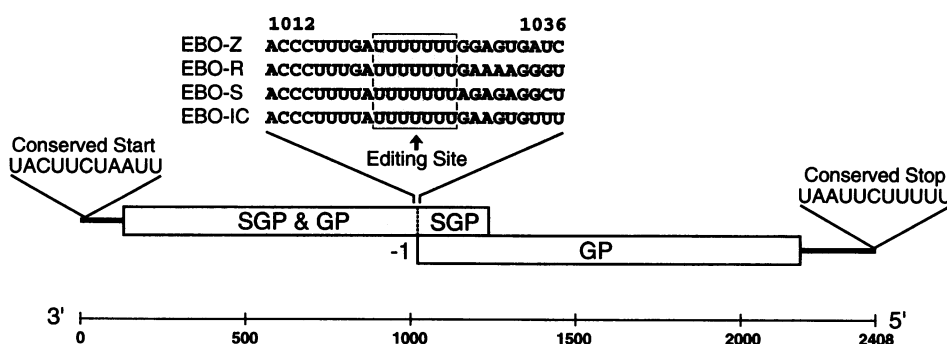


FIG. 1. Schematic representation of the GP gene organization for the EBO-Z, EBO-R, EBO-S, and EBO-IC subtypes. GenBank accession numbers for EBO virus sequences are U23187 (Yambuku/1976/057935), U28077 (Kikwit/1995), U23152 (Reston/1989), U23416 (Manila/1992; ORF only), U23417 (Siena/1992; ORF only), U28134 (Maridi/1976), U23069 (Nzara/1979), and U28006 (Tai Forest/1994). Shown are the conserved transcriptional start and stop signals, coding regions, and the site of transcriptional editing. An alignment of the region containing the transcriptional editing sites for the four subtypes of EBO virus is also shown (EBO-Z sequence shown = 1012–1036). All sequences are shown in the minus sense. The scale at the bottom represents the length of the EBO-Z GP gene (2408 nucleotides).

region (Fig. 2) where more than half of the potential N-linked glycosylation sites and the putative region for O-linked glycosylation are located. The rest of the GP sequence is conserved and is more hydrophobic. In addition, all the cysteine residues in the SGP and GP are conserved (see asterisks in Fig. 2). One common feature found in the GPs of all filoviruses is a highly conserved immunosuppressive motif (26 residues) in the C-terminal third of the GP that has a high degree of homology to those found in the envelope glycoproteins of oncogenic retroviruses (9, 18).

Since sequence analysis indicated that the GPs of EBO viruses are encoded in two frames, it was important to determine the mechanism(s) by which the GP is expressed. The original published description of the EBO-Z (Yambuku/1976/057935) GP gene (7) was based on a virus stock that had undergone three rounds of plaque purification and was later determined to contain a single base insertion (extra uridine nucleotide) that joined the GP frames. Because a single base insertion led to the connecting of the frames in the mutant (plaque purified) EBO-Z, it was thought that RNA editing at or near this site, in a manner similar to that described for the P genes of paramyxoviruses (19), may be responsible for the generation of GP. This region of the EBO-Z (Yambuku/1976/057935) and EBO-R (Reston/1989) GP mRNA were RT-PCR amplified and cloned, and the isolated clones were sequenced. From a total of 81 EBO-R and 41 EBO-Z clones, an addition of a single nontemplated adenosine in the mRNA at the suspected site was found in 13 (16%) and 11 (27%) of the clones, respectively. Fig. 1 shows a short alignment of the GP gene region containing the editing site for the four subtypes of EBO virus. To eliminate the possibility that the insertion of an extra adenosine merely reflects template (vRNA) variability, RT-PCR was directed at negative sense EBO-Z (Kikwit/1995) RNA isolated from a human blood sample. Sequencing of 32

clones showed no insertions in the template vRNA at the editing site, indicating that the genomic sequence is homogeneous and transcriptional editing is responsible for the heterogeneity seen in the GP mRNA clones.

To confirm that the synthesis of SGP and GP are directed by unedited and edited transcripts, respectively, *in vitro* expression of the EBO-R (Reston/1989) GP ORF was performed. Two plasmids were used to generate run-off transcripts, one that contained seven adenosines at the editing site (mRNA sense) and one that had eight adenosines (GP frames connected). Products of *in vitro* translation of run-off transcripts are seen in Fig. 3A. Translation of the RNA containing seven adenosines (unedited form) resulted in the synthesis of an unglycosylated protein that is close to the predicted size of the SGP backbone (41.8 kDa) (Fig. 3A, lane 1). This protein was processed to a higher molecular weight when the transcript was translated in the presence of canine pancreatic microsomal membranes (lane 2). Similarly, the RNA with eight adenosines (edited form) produced a prominent protein in the predicted size range for the unglycosylated GP and could also be processed to a higher molecular weight form (lanes 3 and 4). Translation of the transcript with seven adenosines produced not only SGP, but also a small amount of unglycosylated GP that was processed in the presence of membranes. Synthesis of GP from the unedited transcript may have occurred via translational frameshifting in a -1 direction (20, 21), or may have occurred due to addition of an extra nucleotide at the editing site (or some other site) by the SP6 RNA polymerase, similar to the transcriptional editing carried out by the EBO virus polymerase. The processed forms of SGP and GP expressed *in vitro* correspond in size to SGP and GP immunoprecipitated from the cells and TCF of EBO-R-infected MA-104 cultures (Fig. 3A, lanes 5–8). The SGP produced *in*

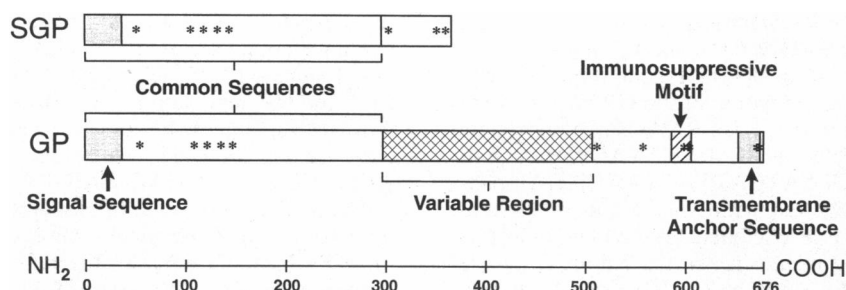


FIG. 2. Profiles of SGP and GP amino acid sequences of EBO subtypes. SGP and GP share the same N-terminal ≈ 300 residues. Identified on the schematic drawing are the signal sequences for the SGP and GP, the GP C-terminal transmembrane anchor sequence, and an immunosuppressive motif found in the GP. Conserved cysteine residues are identified by asterisks. A highly variable central region of the GP is identified, while all other regions are relatively conserved. The lengths of the SGP and GP for the EBO-Z subtype are 364 and 676 amino acids, respectively.

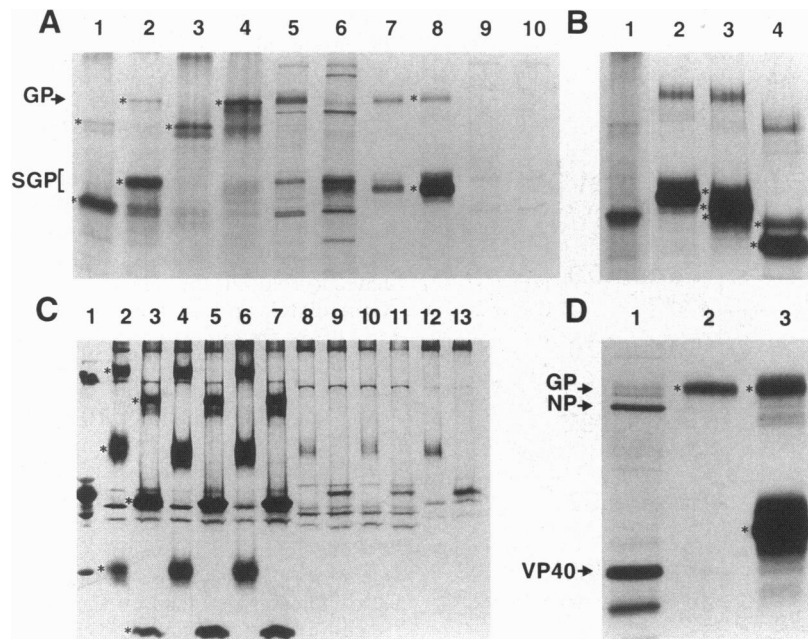


FIG. 3. Expression of EBO virus GPs. (A) Fluorogram of [35 S]methionine/cysteine-labeled EBO-R GPs expressed *in vitro* and *in vivo* separated on SDS/polyacrylamide gels (10% gel). *In vitro* translation products of run-off transcripts generated from the plasmid pSP73-RESGP-7A (lanes 1 and 2) and pSP73-RESGP-8A (lanes 3 and 4), each pair without and with addition of canine pancreas membranes, respectively. Asterisks next to lanes 1–4 identify the unprocessed and larger processed (glycosylated) forms of SGP and GP. The remaining lanes contain RIP products (bound by an anti-EBO-R GP/SGP monoclonal antibody) from EBO-R (Reston/1989)-infected MA-104 cell lysates (lanes 5 and 6), supernatant fluids (lanes 7 and 8), or mock-infected cell supernatant fluids (lanes 9 and 10). Lanes 5, 7, and 9 contain products derived from lysates prepared by SDS treatment (followed by Nonidet P-40 and deoxycholate), and lanes 6, 8, and 10 were treated with Triton X-100 (13). At the left edge are marked the positions where GP and the broader SGP bands migrate. Asterisks next to lane 8 identify SGP and GP immunoprecipitated from supernatant fluids. (B) Endoglycosidase treatment of the EBO-R GP and SGP. Lane 1 (marker lane) contains the same *in vitro* translation products as lane 1 in A. The same preparation separated in lane 8 in A was mock-treated (lane 2) or digested with either endoglycosidase H (lane 3) or endoglycosidase F/N-glycosidase F mixture (lane 4). Asterisks identify digestion products. (C) Immunoprecipitation of SGP and GP from EBO virus-infected cell supernatant fluids. Lane 1 (marker lane) contains EBO-Z (Yambuku/1976/057935) virion proteins. The remaining lanes are paired (2 and 3, 4 and 5, etc.) and contain mock-treated (even lanes) and endoglycosidase F/N-glycosidase F-treated (odd lanes) proteins immunoprecipitated with a mouse polyclonal anti-EBO-Z GP. RIP products were derived from the supernatant fluids of MA-104 cells infected with three different EBO-Z (lanes 2–7; viruses Yambuku/1976/057935 and 057878, and Tandala/1977, respectively), two EBO-S (lanes 8–11; viruses Nzara/1979 and Maridi/1976), and EBO-R (lanes 12 and 13; Reston/1989). Asterisks identify three prominent GP bands and deglycosylated products seen in EBO-Z lanes. (D) Demonstration that the SGP of EBO-R is a nonstructural GP. Purified virion preparations, labeled with either [35 S]methionine (lane 1) or [3 H]glucosamine (lane 2), are shown together with the same RIP products as lane 8 in A. Asterisks identify the virion GP in lanes 2, and the GP and SGP in lane 3. At the left edge are marked the locations of the virion GP, nucleoprotein (NP), and VP40 (matrix protein).

in vitro, however, did not appear as broad (heterogeneous) as the SGP detected in the TCF of EBO-R-infected cultures.

The secreted EBO-R SGP ranges in size from 50 to 70 kDa, with the central major band estimated to be 59 kDa. Endoglycosidase digestion indicated that at least three major species are present (Fig. 3B). Removal of all N-linked glycans by Endo F/NF produced two bands, a smaller predominant species and a weaker larger band that may contain O-linked glycans or represents some other form of the SGP. The SGP found in the supernatant fluids of cells infected with EBO-Z and EBO-S are comparable with those seen with EBO-R (Fig. 3C). The EBO-Z strains showed strong bands that correspond to the GP and SGP described above for EBO-R, but cross-reactivity of the anti-EBO-Z GP with the EBO-S and EBO-R subtypes resulted in much weaker and thinner bands. Surprisingly, the Zaire subtypes produced a third glycosylated protein band (≈ 24 kDa) of unknown origin. Fig. 3D shows that the SGP of EBO-R is absent from purified virion preparations (as is the case for the other EBO subtypes) and is thus a nonstructural protein. As with the SGP, the smallest GP noted above has not been detected in virions.

The EBO virus GP genes differed from those of MBG virus by at least 55% at the nucleotide level and 67% at the amino acid level. Even among the EBO subtypes a high degree of genetic variability was evident. Four clearly distinct EBO virus subtypes (Zaire, Sudan, Reston, and Ivory Coast) were iden-

tified, differing from one another by 37–41% and 34–43% at the nucleotide and amino acid level, respectively. A significant finding was that the nucleotide sequence of the Kikwit/1995 EBO-Z differed from the Yambuku/1976/057935 isolate by <1.6% over the entire gene. In addition, no amino acid sequence differences are predicted for the SGPs of these viruses, but a 2% difference in the GP was noted, primarily in the variable region where 11 of the 13 amino acid changes occurred.

A detailed phylogenetic profile of the family Filoviridae was determined using nucleotide sequences that encode N- and C-terminal regions of the GP. These regions contained sufficient homology to allow accurate nucleotide alignment. Maximum parsimony analysis of these sequences produced a single most parsimonious tree that clearly separated EBO and MBG viruses and divided the EBO subtypes into four distinct clades (Fig. 4).

DISCUSSION

The finding that all EBO subtypes encode their structural (virion) GPs in two frames and that expression of these GPs occurs through transcriptional editing is highly unusual. Our results indicated that there was no sequence variation in the EBO vRNA corresponding to the mRNA editing site when analyzed directly from the blood of an acutely infected human.

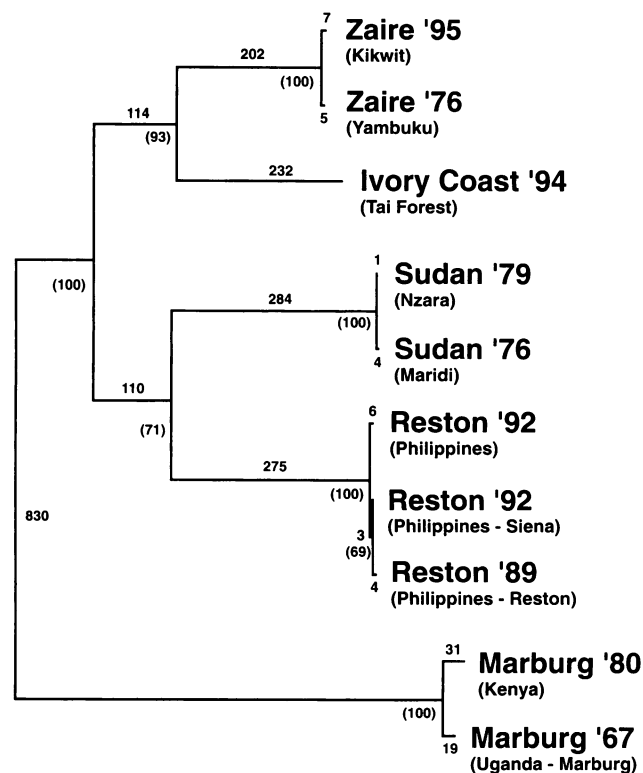


FIG. 4. Phylogenetic relationship of filovirus GP gene sequences. Shown is a phylogenetic tree obtained from an alignment of concatenated nucleotide sequences that encode conserved N-terminal and C-terminal regions of filovirus GP ORFs. The ORF sequences for EBO-Z, EBO-IC, and EBO-S are nucleotides 46–536 + 1528–1885, for EBO-R 49–539 + 1531–1888, and for MBG 2–489 + 1532–1589 (1 = start of ORF). Maximum parsimony analysis (branch-and-bound with a 4:1 weighting of transversions to transitions) produced a single most parsimonious tree. The values for bootstrap confidence limits (%) for branch points were generated from 1000 replicates of the analysis and are indicated in parentheses at branch points. GenBank accession numbers for sequences used in phylogenetic analyses are U28077, U23187, U28006, U23069, U28134, U23416, U23417, U23152, Z12132, and Z29337.

This reinforces our belief that the encoding of the GP in two frames may be important in the maintenance of EBO in nature and perhaps also in the pathogenesis of human disease. We have no data bearing on the consequences of the circulation of a soluble protein bearing the N-terminal 300 amino acids of the virion GP, and we have not found homologues of the novel SGP sequences (C-terminal 70–77 residues) that suggest a specific role in the infectious process. We speculate that the SGP may interact with the immune system, either at the cellular level, where the result could be cellular deletion, energy, or activation of suppression, or at the humoral effector level, where high-affinity antibodies directed against the N-terminal portion of the GP may be prevented from acting on virions or cells presenting surface GP. Indeed, patients usually die without evidence of an effective immune response, and even when recovery is underway survivors do not have detectable virus neutralizing antibodies. Our preliminary investigations have shown that SGP is detectable in high concentrations in the blood of acutely infected patients (data not shown). If the SGP does play an important role in the pathogenesis of human infection, it may be possible to protect or increase survival rates if it can be targeted by specific antibodies early in the infection.

The discovery of a third and even smaller 24-kDa glycoprotein in the medium of EBO-Z-infected cells was unexpected. This protein, which is reactive with an anti-EBO-Z GP serum,

may result from cleavage or degradation of GP or SGP or may be produced as a result of a premature termination event during translation. Further studies are required in order to determine how this protein is produced in infected cells and if this protein is unique to certain EBO subtypes.

The immunosuppressive motif identified in the C-terminal third of filovirus GPs is the most conserved sequence seen between the GPs of EBO and MBG viruses (7), and may contribute to the immunosuppression observed in humans and monkeys infected with filoviruses. In addition, this motif may have a role in the assembly of the GP into peplomers. Immediately upstream and slightly overlapping the N-terminal end of the immunosuppressive motifs of all filovirus GPs is a region that contains a heptad repeat sequence that Chambers *et al.* (22) identified in the fusion GPs of oncogenic retroviruses (the same location with respect to the immunosuppressive motif of p15E) and paramyxoviruses. They postulated that these repeats may form an extended backbone for the spike structures through the coiling of the GPs around one another at these regions.

Phylogenetic analysis of GP gene nucleotide sequences clearly distinguishes the EBO subtypes from MBG virus, but also indicates that each of the four EBO subtypes represents a monophyletic lineage. The close phylogenetic relationship of EBO-R strains, filoviruses associated with monkeys exported from the Philippines (23–25), to the African subtypes may indicate that this EBO-R is not indigenous to Asia but may have been introduced to this region from Africa.

The similarity in the GP genes of the 1976 and 1995 EBO-Z strains (<1.6% for the entire gene) is surprising, since more than 18 years and 1000 km separate the outbreaks caused by these viruses. These data may indicate that EBO viruses (and filoviruses in general) have evolved to occupy relatively stable specific niches in nature and may not change appreciably in their natural hosts. The high degree of similarity between the 1976 and 1995 EBO-Z viruses suggests that the reservoir is the same in both locations and that it is widespread in Zaire or is a migratory species. In contrast, the divergence seen among the different EBO subtypes and between EBO and MBG viruses might also imply that these viruses are very old and have slowly coevolved with their as-yet-unknown natural hosts, which may also demonstrate a similar extent of genetic variability. While the genetic diversity among the four EBO virus subtypes is high, the apparent stability of each subtype bodes well for the development of effective immune or antiviral therapy or vaccination strategies. A large scale effort is currently underway to attempt to discover the natural reservoir for EBO viruses. This includes screening of large numbers of vertebrates and invertebrates from the Kikwit area for evidence of EBO virus infection. Currently, we have insufficient knowledge to help narrow the focus of this search, although considerable effort is being placed on vertebrates in the collection. This decision is influenced by the genetic stasis observed between the EBO-Z strains from 1976 and 1995, and by the fact that all filoviruses examined possess an immunosuppressive domain in their GP that to date has only been found in vertebrate RNA viruses.

In conclusion, we have described an unusual type of organization for the GP genes of all EBO viruses, which has not been described for any other type of virus gene encoding a structural GP. We feel that the expression of this gene is an important element in the pathogenesis of EBO viruses and is critical to their maintenance in the natural host. In the future we plan to continue studying the expression of EBO virus GPs *in vitro* and *in vivo*, to examine the roles of SGP and GP in the disease process, and to determine if these proteins can serve as protective immunogens.

We wish to thank A. L. Moen and L. Brammer for their technical assistance in nucleotide sequencing and hybridoma production, T. G.

Ksiazek and P. E. Rollin for their laboratory support in isolating and providing stocks of the Kikwit/1995 EBO-Z virus, and J. P. O'Connor for editing this work. We also thank T. Muyembe and G. Van der Groen for providing us with human specimens from Zaire, and B. Le Guenno and D. Brown for supplying the EBO-IC (Tai Forest/1994) and EBO-R (Sienna/1992) isolates.

- Centers for Disease Control and Prevention (1995) *Morbid. Mortal. Week. Rep.* **44**, 381–382.
- Le Guenno, B., Formenty, P., Wyers, M., Gounon, P., Walker, F. & Boesch, C. (1995) *Lancet* **345**, 1271–1274.
- Feldmann, H., Klenk, H.-D. & Sanchez, A. (1993) *Arch. Virol. Suppl.* **7**, 81–100.
- Peters, C. J., Sanchez, A., Rollin, P. E., Ksiazek, T. G. & Murphy, F. A. (1995) in *Virology*, ed. Fields, B. N. (Raven, New York), 3rd Ed., in press.
- Peters, C. J., Sanchez, A., Feldmann, H., Rollin, P. E., Nichol, S. & Ksiazek, T. G. (1994) *Semin. Virol.* **5**, 147–154.
- Sanchez, A., Ksiazek, T. G., Rollin, P. E., Peters, C. J., Nichol, S. T., Khan, A. S. & Mahy, B. W. J. (1995) *Emerging Inf. Dis.* **1**, 96–97.
- Sanchez, A., Kiley, M. P., Holloway, B. P. & Auperin, D. D. (1993) *Virus Res.* **29**, 215–240.
- Geyer, H., Will, C., Feldmann, H., Klenk, H.-D. & Geyer, R. (1992) *Glycobiology* **2**, 299–312.
- Will, C., Mülberger, E., Linder, D., Slenczka, W., Klenk, H.-D. & Feldmann, H. (1993) *J. Virol.* **67**, 1203–1210.
- Feldmann, H., Nichol, S. T., Klenk, H.-D., Peters, C. J. & Sanchez, A. (1994) *Virology* **199**, 469–473.
- McCormick, J. B., Bauer, S. P., Elliott, L. H., Webb, P. A. & Johnson, K. M. (1983) *J. Infect. Dis.* **147**, 264–267.
- Sanchez, A. & Kiley, M. P. (1987) *Virology* **157**, 414–420.
- Gübler, U. & Hoffman, G. J. (1983) *Gene* **25**, 263–269.
- Sanchez, A., Kiley, M. P., Holloway, B. P., McCormick, J. B. & Auperin, D. D. (1989) *Virology* **170**, 81–91.
- Sanger, F., Nicklen, S. & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 5463–5467.
- Sanchez, A., Pifat, D. Y., Kenyon, R. H., Peters, C. J., McCormick, J. B. & Kiley, M. P. (1989) *J. Gen. Virol.* **70**, 1125–1132.
- Sanchez, A. & Frey, T. K. (1991) *Virology* **183**, 636–646.
- Volchkov, V. E., Blinov, V. M. & Netesov, S. V. (1992) *FEBS Lett.* **305**, 181–184.
- Lamb, R. A. & Paterson, R. G. (1993) in *The Paramyxoviruses*, ed. Kingsbury, D. W. (Plenum, New York), pp. 181–214.
- Gallant, J. & Lindsley, D. (1993) *Biochem. Soc. Trans.* **21**, 817–821.
- Brierley, I. (1995) *J. Gen. Virol.* **76**, 1885–1892.
- Chambers, P., Pringle, C. R. & Easton, A. J. (1990) *J. Gen. Virol.* **71**, 3075–3080.
- Jahrling, P. B., Geisbert, T. W., Dalgard, D. W., Johnson, E. D., Ksiazek, T. G., Hall, W. C. & Peters, C. J. (1990) *Lancet* **335**, 502–505.
- Centers for Disease Control and Prevention (1990) *Morbid. Mortal. Week. Rep.* **38**, 831–838.
- World Health Organization (1992) *WER* **67**, 142–143.